


 Europäisches Patentamt  
 European Patent Office  
 Office européen des brevets

⑪ Publication number:

0 079 739  
 A2

⑫

# EUROPEAN PATENT APPLICATION

⑲ Application number: 82305926.6

⑳ Date of filing: 08.11.82

⑤① Int. Cl.<sup>3</sup>: C 12 N 15/00  
 C 12 N 1/00, C 12 P 21/02  
 C 07 H 21/04, C 07 C 103/52  
 //C12R1/19, C12R1/865

③① Priority: 12.11.81 US 320632

④③ Date of publication of application:  
 25.05.83 Bulletin 83/21

⑥④ Designated Contracting States:  
 BE CH DE FR GB IT LI NL SE

⑦① Applicant: THE UPJOHN COMPANY  
 301 Henrietta Street  
 Kalamazoo, Michigan 49001(US)

⑦② Inventor: Dugaiczky, Achilles  
 c/o The Upjohn Company 301 Henrietta Street  
 Kalamazoo Michigan 49001(US)

⑦④ Representative: Perry, Robert Edward et al,  
 GILL JENNINGS & EVERY 53-64 Chancery Lane  
 London WC2A 1HN(GB)

⑤④ Albumin-based nucleotides, their replication and use, and plasmids for use therein.

⑤⑦ The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A2

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION  
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly  $\alpha$ -fetoprotein, but the synthesis decreases drastically after birth. Recently, 10 Law et al determined the complete sequence of mouse  $\alpha$ -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the  $\alpha$ -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched  
 5 albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter  
 10 was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending  
 15 into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by  
 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a  
 25 prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table  
 30 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T<sup>C</sup> T C T T C T G T.....albumin mRNA  
 35 (3')...G A G G A A G G C G U C C m<sub>2</sub><sup>6</sup>A m<sub>2</sub><sup>6</sup>A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre-peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the pro-peptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

		-18	p r o	-10	
		Met lys trp val tlu phe lle ser leu leu phe leu phe ser			
		GCTTTTCTCTCTGTCACCCACACCCTTTGCCACA ATG AAG TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT AGC			(30)
		-1	p r o	-1	
		ser ala tyr ser arg gly val phe arg asp ala his lys ser glu val ala his arg phe lys asp leu qly glu asn phe lys			
		TCC GCT TAT TCC ACC GGT GTG TTT CGT CGA GAT GCA CAC AAC AGT CAG GCT CAT CGC TTT AAA GAT TTG GGA GAA AAT TTC AAA			(170)
		21	30	34	40
		ala leu val leu ile ala phe ala gln tyr leu gln cys pro phe glu asp his val lys leu val asn glu val thr qlu phe ala			
		GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA			(260)
		51	53	60	62
		lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu			
		AAA ACA TGT GCT GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA TCA CTT CAT ACC CTT TTT CGA CAC AAA TTA TGC ACA GTT GCA ACT CTT			(350)
		81	90	91	100
		arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro qly arg asn qlu cys phe leu qln his lys asp asp asn pro			
		CGGT GAA ACC TAT CGT GAA ATG GCT CAC TGC TGT GCA AAA CAA GAA CCT GCG ACA AAT GAA TGC TTC CAA CAC AAA GAT GAC AAC CCA			(440)
		111	120	124	130
		asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu qlu thr phe leu lys lys tyr leu try			
		AAC CTC CCC CGA TTG GTG ACA CCA CAG GTT CAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT			(330)
		141	150	160	168
		glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr qlu cys cys qln			
		GAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCC GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA GAA TGT TGC CAA			(620)
		171	177	180	190
		ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp qlu qly lys ala ser ala lys qln arg leu lys cys			
		GCT GCT GAT AAA GCT GCC TGC CTG TTG CCA AAG CTC GAT GAA CCG MAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT			(710)
		201	210	220	230
		ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser qln arg phe pro lys ala ala phe ala qlu			
		GCC AGT CTC CAA AAA TTT CGA GAA ACA GCT TTC AAA GCA TGG GCA GTA GCT CGC CTG AGC CAG ACA TTT CCC AAA CCT GAG TTT GCA GAA			(300)

231 val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TGC TGC CAT GCA GAT CTG CTT GAA TGT GCT CAT CAC ACG GCG GAC CTT (890)

240 245 246 250 253 260  
 261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile  
 CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TCC ATT (980)

270 278 279 280 289 290  
 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
 GCC GAA GTG GAA AAT GAT GAG ATG CCT CCT GAC TTG CCT TCA TTA GCT GCT CAT TTT GTT GAA AGT AAG CAT GTT TGC AAA AAC TAT CTT (1070)

300 310 316 320  
 321 glu ala lys asp val phe leu glu met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala  
 CAG GCA AAG GAT GTC TTC TTG GCG ATG TTT TTG TAT GAA TAT GCA AGA AGC CAT CCT GAT TAC TCT GTC CTG CTG CTG ACA CTT GCC (1160)

330 340 350  
 351 lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TCC TGT GCT GCT GCA GAT CCT CAT GAA TGC TAT GCG AAA GTG TTC GAT GAA TTT AAA CCT CTT (1250)

360 361 369 370 380  
 381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu leu qly qly tyr lys phe qln asn ala leu leu val arg  
 GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CTT TTT GAG CAG CTT GGA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CGT (1340)

390 392 400 410  
 411 thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA AGA AAC CTA CCA AAA GTG GCG AGC AAA TGT TGT AAA CAT (1430)

420 430 437 438 440  
 441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser  
 CCT GAA GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC CTC AAC CAG TTA TGT GTG TTG CAT CAG AAA ACG CCA GTA AGT (1520)

448 450 460 461 470  
 471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
 GAC AGA GTC ACC AAA TGC TGC ACA GAA TCC TTG GTG AAC ACG CCA TCC TTT TCA GCT CTG CAA GTC CAT GAA ACA TAC GTT CCC AAA (1610)

476 477 480 490 500  
 501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys qly arg qln ile lys lys qln thr ala leu val  
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AAG GAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

510 514 520 530

531 540 550 558 559 560  
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys cys lys  
GAG CTC GTG AAA CAC CAC AAG CAC CCA ACA AAA GAG CAA CTG AAA CCT GTT ATG GAT CAT TTC GCT GCT TTT GTA GAG AAG TCC TGF AAG (1790)  
561 567 570 580  
ala asp asp lys glu thr cys phe ala glu gln qly lys leu val ala ala ser gln ala ala leu qly leu ter  
CCT GAC GAT AAG GAG ACC TCC TTT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA CCT GCC TTA TAA CATCACATTTAAAG (1883)  
ter ter  
CATCTCAGCCTACCATGAGAAATGAGCAAGAAAATGACATCAAGCCTTATTCATCTGTTTCTTTTCTGCTGTAAAGCCCAACCCCTGCTCTAAACATAAATTTCTTAA (2002)  
TCATTTGCCCTTTTCTGCTGCTTCAATTAATAAATAATGCAACATCTAA..... 20 .....AA (2078)

Following are examples which illustrate procedures, ~~including the best mode~~ for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5    Example 1            Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685].

15    Example 2            Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [<sup>32</sup>P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30    As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HB101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to  
35    separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris-HCl, pH 8.0, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>). The cells for transformation are



prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3      Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

20 Example 4      DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and  $\gamma$ [<sup>32</sup>P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

30 Example 5      Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5 HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public, ~~upon the grant of a patent. It should be understood that the availability~~  
10 of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject ~~instrument by governmental action.~~

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL  
15 B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEpl6 is a well known and widely available yeast episomal plasmid.  
20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEpl6) - NRRL B-12093.

Example 6      Assembly of the Serum Albumin Gene

Assembling the pieces together is a straightforward task of re-  
25 striction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the  
30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

- 35 (a) Escherichia coli  
(b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRI DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco RI cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7      Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Pujalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

25 Example 8      Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

35

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies  
5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T.  
10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number  
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number  
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

35  
 30  
 25  
 20  
 15  
 10  
 5  
 -10  
 -1 -6 -1 -1 -1  
 ser ala tyr ser arg gly val phe arg asp ala his lys ser glu val ala his arg phe lys asp leu aly glu asn phe lys  
 TCG GCT TAT TCC ACG GGT GTG TTT CGT CGA GAT GCA CAC AAC AGT CAG GTT GCT CAT CGC TTT AAA GAT TTG GCA GAA CAA AAT TTC AAA (170)  
 21  
 ala leu val leu lle ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala  
 GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)  
 51  
 lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu  
 AAA ACA TGT GTT GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CCA GAC AAA TTA TGC ACA GTT GCA ACT CTT (350)  
 81  
 arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu gln his lys asp asp asn pro  
 CGT GAA ACC TAT GGT GAA ATG GCT GAC TGC TGT GCA AAA CAA GAA CCA CCT GCG ACA AAT GAA TGC TTC TTG CAA CAC AAA GAT GAC AAC CCA (440)  
 111  
 asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try  
 AAC CTC CCC CCA TTG GTG ACA CCA GAG GTT GAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT (530)  
 141  
 glu lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala phe thr glu cys cys gln  
 GAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA GAA TGT TGC CAA (620)  
 171  
 ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg leu lys cys  
 GCT GCT CAT AAA GCT GCC TGC CTG TTG CCA AAG CTC GAT GAA CTT CCG GAT GAA GCG MAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (710)  
 201  
 ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu  
 CCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA GCA TGG GCA GTA GCT CCC CTG AGC CAG ACA TTT CCC AAA GCT GAG TTT GCA CAA (800)

231 val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu 260  
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACC GAA TGC TGC CAT GCA GAT CTG CTT GAA TGT GCT GAT CAC ACC GCG GAC CTT (890)  
 261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys oys qiu lys pro leu leu glu lys ser his cys ile 289 290  
 CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TCC TGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TCC ATT (980)  
 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val oys lys asn tyr ala 320  
 GCC GAA GTG GAA AAT GAT CAG ATG CCT GCT CAC TTG CCT TCA TTA GCT GCT CAT TTT GTT GAA AGT AAG GAT GTT TGC AAA AAC TAT GCT (1070)  
 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350  
 CAG GCA AAG GAT GTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA ACA AGG CAT CCT GAT TAC TCT GTC CTG CTG CTG ACA CTT GCC (1160)  
 351 lys thr tyr glu thr thr leu glu lys oys cys ala ala asp pro his qiu cys tyr ala lys val phe asp glu phe lys pro leu 380  
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TCC TGT TGT GCT GCT GCT GCA GAT CCT CAT GAA TCC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)  
 381 val glu glu pro gln asn leu ile lys gln asn cys qiu leu phe qiu leu qly glu tyr lys phe gln asn ala leu leu val arg 410  
 GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CAG CTT TTT GAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CGT (1340)  
 411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA GCA AAA GTG GCG AGC AAA TGT TGT AAA CAT (1430)  
 441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu oys val leu his qiu lys thr pro val ser 470  
 CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAT CTA TCC GTC GTC CAC TAT CTA TCC GTC GTC AAC CAG TTA TGT GTG TGC CAT GAG AAA ACC CCA GTA AGT (1520)  
 471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp qiu thr tyr val pro lys 500  
 CAC ACA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG CGA CCA TCC TTT TCA GCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)  
 501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser qiu lys glu arg qin ile lys lys qin thr ala leu val 530  
 CAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT CAG AAG CAG ACA CAA AAT CAA ACT GCA CTT GCT (1700)

5  
10  
15  
20  
25  
30  
35

531 558 559 560  
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys lys  
GAG CTC GTG AAA CAC AAG CCC AAG GCA ACA AAA GAG CAA CTG AAA CCT GTT ATG CAT CAT TTC GCT TTT GTA GAG AAG TCC TGC AAG (1790)

561 567 570 580  
ala asp asp lys glu thr cys phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter  
GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CTT GGT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAG (1883)

ter ter  
CATCTCAGCCTACCATGACATTAAGAGAAAGAAATGAAGATCAAAAGCCTTATTCATCTGTGTTTCTTTTTCGTTGGTGTAAAGCCACCCCTGTCTTAAACATAAATTTCCTTAA (2002)

TCATTTGCCCTCTTTTCTCTGCTTCAATTATAAATAATGGAAGAAATCTAA..... 20 .....AA (2078)



6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

```

5
10
15
20
25
30
35
1
asp ala his lys ser glu val ala his arg phe lys asp leu gln glu asn phe lys 20
GAT CCA CAC AAG AGT GAG GTT CAT CGG TTT AAA GAT TTG CCA GAA GAA AAT TTC AAA (170)
10
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
1440
1450
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720
1730
1740
1750
1760
1770
1780
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100
2110
2120
2130
2140
2150
2160
2170
2180
2190
2200
2210
2220
2230
2240
2250
2260
2270
2280
2290
2300
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2410
2420
2430
2440
2450
2460
2470
2480
2490
2500
2510
2520
2530
2540
2550
2560
2570
2580
2590
2600
2610
2620
2630
2640
2650
2660
2670
2680
2690
2700
2710
2720
2730
2740
2750
2760
2770
2780
2790
2800
2810
2820
2830
2840
2850
2860
2870
2880
2890
2900
2910
2920
2930
2940
2950
2960
2970
2980
2990
3000
3010
3020
3030
3040
3050
3060
3070
3080
3090
3100
3110
3120
3130
3140
3150
3160
3170
3180
3190
3200
3210
3220
3230
3240
3250
3260
3270
3280
3290
3300
3310
3320
3330
3340
3350
3360
3370
3380
3390
3400
3410
3420
3430
3440
3450
3460
3470
3480
3490
3500
3510
3520
3530
3540
3550
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760
3770
3780
3790
3800
3810
3820
3830
3840
3850
3860
3870
3880
3890
3900
3910
3920
3930
3940
3950
3960
3970
3980
3990
4000
4010
4020
4030
4040
4050
4060
4070
4080
4090
4100
4110
4120
4130
4140
4150
4160
4170
4180
4190
4200
4210
4220
4230
4240
4250
4260
4270
4280
4290
4300
4310
4320
4330
4340
4350
4360
4370
4380
4390
4400
4410
4420
4430
4440
4450
4460
4470
4480
4490
4500
4510
4520
4530
4540
4550
4560
4570
4580
4590
4600
4610
4620
4630
4640
4650
4660
4670
4680
4690
4700
4710
4720
4730
4740
4750
4760
4770
4780
4790
4800
4810
4820
4830
4840
4850
4860
4870
4880
4890
4900
4910
4920
4930
4940
4950
4960
4970
4980
4990
5000
5010
5020
5030
5040
5050
5060
5070
5080
5090
5100
5110
5120
5130
5140
5150
5160
5170
5180
5190
5200
5210
5220
5230
5240
5250
5260
5270
5280
5290
5300
5310
5320
5330
5340
5350
5360
5370
5380
5390
5400
5410
5420
5430
5440
5450
5460
5470
5480
5490
5500
5510
5520
5530
5540
5550
5560
5570
5580
5590
5600
5610
5620
5630
5640
5650
5660
5670
5680
5690
5700
5710
5720
5730
5740
5750
5760
5770
5780
5790
5800
5810
5820
5830
5840
5850
5860
5870
5880
5890
5900
5910
5920
5930
5940
5950
5960
5970
5980
5990
6000
6010
6020
6030
6040
6050
6060
6070
6080
6090
6100
6110
6120
6130
6140
6150
6160
6170
6180
6190
6200
6210
6220
6230
6240
6250
6260
6270
6280
6290
6300
6310
6320
6330
6340
6350
6360
6370
6380
6390
6400
6410
6420
6430
6440
6450
6460
6470
6480
6490
6500
6510
6520
6530
6540
6550
6560
6570
6580
6590
6600
6610
6620
6630
6640
6650
6660
6670
6680
6690
6700
6710
6720
6730
6740
6750
6760
6770
6780
6790
6800
6810
6820
6830
6840
6850
6860
6870
6880
6890
6900
6910
6920
6930
6940
6950
6960
6970
6980
6990
7000
7010
7020
7030
7040
7050
7060
7070
7080
7090
7100
7110
7120
7130
7140
7150
7160
7170
7180
7190
7200
7210
7220
7230
7240
7250
7260
7270
7280
7290
7300
7310
7320
7330
7340
7350
7360
7370
7380
7390
7400
7410
7420
7430
7440
7450
7460
7470
7480
7490
7500
7510
7520
7530
7540
7550
7560
7570
7580
7590
7600
7610
7620
7630
7640
7650
7660
7670
7680
7690
7700
7710
7720
7730
7740
7750
7760
7770
7780
7790
7800
7810
7820
7830
7840
7850
7860
7870
7880
7890
7900
7910
7920
7930
7940
7950
7960
7970
7980
7990
8000
8010
8020
8030
8040
8050
8060
8070
8080
8090
8100
8110
8120
8130
8140
8150
8160
8170
8180
8190
8200
8210
8220
8230
8240
8250
8260
8270
8280
8290
8300
8310
8320
8330
8340
8350
8360
8370
8380
8390
8400
8410
8420
8430
8440
8450
8460
8470
8480
8490
8500
8510
8520
8530
8540
8550
8560
8570
8580
8590
8600
8610
8620
8630
8640
8650
8660
8670
8680
8690
8700
8710
8720
8730
8740
8750
8760
8770
8780
8790
8800
8810
8820
8830
8840
8850
8860
8870
8880
8890
8900
8910
8920
8930
8940
8950
8960
8970
8980
8990
9000
9010
9020
9030
9040
9050
9060
9070
9080
9090
9100
9110
9120
9130
9140
9150
9160
9170
9180
9190
9200
9210
9220
9230
9240
9250
9260
9270
9280
9290
9300
9310
9320
9330
9340
9350
9360
9370
9380
9390
9400
9410
9420
9430
9440
9450
9460
9470
9480
9490
9500
9510
9520
9530
9540
9550
9560
9570
9580
9590
9600
9610
9620
9630
9640
9650
9660
9670
9680
9690
9700
9710
9720
9730
9740
9750
9760
9770
9780
9790
9800
9810
9820
9830
9840
9850
9860
9870
9880
9890
9900
9910
9920
9930
9940
9950
9960
9970
9980
9990
10000
10010
10020
10030
10040
10050
10060
10070
10080
10090
10100
10110
10120
10130
10140
10150
10160
10170
10180
10190
10200
10210
10220
10230
10240
10250
10260
10270
10280
10290
10300
10310
10320
10330
10340
10350
10360
10370
10380
10390
10400
10410
10420
10430
10440
10450
10460
10470
10480
10490
10500
10510
10520
10530
10540
10550
10560
10570
10580
10590
10600
10610
10620
10630
10640
10650
10660
10670
10680
10690
10700
10710
10720
10730
10740
10750
10760
10770
10780
10790
10800
10810
10820
10830
10840
10850
10860
10870
10880
10890
10900
10910
10920
10930
10940
10950
10960
10970
10980
10990
11000
11010
11020
11030
11040
11050
11060
11070
11080
11090
11100
11110
11120
11130
11140
11150
11160
11170
11180
11190
11200
11210
11220
11230
11240
11250
11260
11270
11280
11290
11300
11310
11320
11330
11340
11350
11360
11370
11380
11390
11400
11410
11420
11430
11440
11450
11460
11470
11480
11490
11500
11510
11520
11530
11540
11550
11560
11570
11580
11590
11600
11610
11620
11630
11640
11650
11660
11670
11680
11690
11700
11710
11720
11730
11740
11750
11760
11770
11780
11790
11800
11810
11820
11830
11840
11850
11860
11870
11880
11890
11900
11910
11920
11930
11940
11950
11960
11970
11980
11990
12000
12010
12020
12030
12040
12050
12060
12070
12080
12090
12100
12110
12120
12130
12140
12150
12160
12170
12180
12190
12200
12210
12220
12230
12240
12250
12260
12270
12280
12290
12300
12310
12320
12330
12340
12350
12360
12370
12380
12390
12400
12410
12420
12430
12440
12450
12460
12470
12480
12490
12500
12510
12520
12530
12540
12550
12560
12570
12580
12590
12600
12610
12620
12630
12640
12650
12660
12670
12680
12690
12700
12710
12720
12730
12740
12750
12760
12770
12780
12790
12800
12810
12820
12830
12840
12850
12860
12870
12880
12890
12900
12910
12920
12930
12940
12950
12960
12970
12980
12990
13000
13010
13020
13030
13040
13050
13060
13070
13080
13090
13100
13110
13120
13130
13140
13150
13160
13170
13180
13190
13200
13210
13220
13230
13240
13250
13260
13270
13280
13290
13300
13310
13320
13330
13340
13350
13360
13370
13380
13390
13400
13410
13420
13430
13440
13450
13460
13470
13480
13490
13500
13510
13520
13530
13540
13550
13560
13570
13580
13590
13600
13610
13620
13630
13640
13650
13660
13670
13680
13690
13700
13710
13720
13730
13740
13750
13760
13770
13780
13790
13800
13810
13820
13830
13840
13850
13860
13870
13880
13890
13900
13910
13920
13930
13940
13950
13960
13970
13980
13990
14000
14010
14020
14030
14040
14050
14060
14070
14080
14090
14100
14110
14120
14130
14140
14150
14160
14170
14180
14190
14200
14210
14220
14230
14240
14250
14260
14270
14280
14290
14300
14310
14320
14330
14340
14350
14360
14370
14380
14390
14400
14410
14420
14430
14440
14450
14460
14470
14480
14490
14500
14510
14520
14530
14540
14550
14560
14570
14580
14590
14600
14610
14620
14630
14640
14650
14660
14670
14680
14690
14700
14710
14720
14730
14740
14750
14760
14770
14780
14790
14800
14810
14820
14830
14840
14850
14860
14870
14880
14890
14900
14910
14920
14930
14940
14950
14960
14970
14980
14990
15000
15010
15020
15030
15040
15050
15060
15070
15080
15090
15100
15110
15120
15130
15140
15150
15160
15170
15180
15190
15200
15210
15220
15230
15240
15250
15260
15270
15280
15290
15300
15310
15320
15330
15340
15350
15360
15370
15380
15390
15400
15410
15420
15430
15440
15450
15460
15470
15480
15490
15500
15510
15520
15530
15540
15550
15560
15570
15580
15590
15600
15610
15620
15630
15640
15650
15660
15670
15680
15690
15700
15710
15720
15730
15740
15750
15760
15770
15780
15790
15800
15810
15820
15830
15840
15850
15860
15870
15880
15890
15900
15910
15920
15930
15940
15950
15960
15970
15980
15990
16000
16010
16020
16030
16040
16050
16060
16070
16080
16090
16100
16110
16120
16130
16140
16150
16160
16170
16180
16190
16200
16210
16220
16230
16240
16250
16260
16270
16280
16290
16300
16310
16320
16330
16340
16350
16360
16370
16380
16390
16400
16410
16420
16430
16440
16450
16460
16470
16480
16490
16500
16510
16520
16530
16540
16550
16560
16570
16580
16590
16600
16610
16620
16630
16640
16650
16660
16670
16680
16690
16700
16710
16720
16730
16740
16750
16760
16770
16780
16790
16800
16810
16820
16830
16840
16850
16860
16870
16880
16890
16900
16910
16920
16930
16940
16950
16960
16970
16980
16990
17000
17010
17020
17030
17040
17050
17060
17070
17080
17090
17100
17110
17120
17130
17140
17150
17160
17170
17180
17190
17200
17210
17220
17230
17240
17250
17260
17270
17280
17290
17300
17310
17320
17330
17340
17350
17360
17370
17380
17390
17400
17410
17420
17430
17440
17450
17460
17470
17480
17490
17500
17510
17520
17530
17540
17550
17560
17570
17580
17590
17600
17610
17620
17630
17640
17650
17660
17670
17680
17690
17700
17710
17720
17730
17740
17750
17760
17770
17780
17790
17800
17810
17820
17830
17840
17850
17860
17870
17880
17890
17900
17910
17920
17930
17940
17950
17960
17970
17980
17990
18000
18010
18020
18030
18040
18050
18060
18070
18080
18090
18100
18110
18120
18130
18140
18150
18160
18170
18180
18190
18200
18210
18220
18230
18240
18250
18260
18270
18280
18290
18300
18310
18320
18330
18340
18350
18360
18370
18380
18390
18400
18410
18420
18430
18440
18450
18460
18470
18480
18490
18500
18510
18520
18530
18540
18550
18560
18570
18580
18590
18600
18610
18620
18630
18640
18650
18660
18670
18680
18690
18700
18710
18720
18730
18740
18750
18760
18770
18780
18790
18800
18810
18820
18830
18840
18850
18860
18870
18880
18890
18900
18910
18920
18930
18940
18950
18960
18970
18980
18990
19000
19010
19020
19030
19040
19050
19060
19070
19080
19090
19100
19110
19120
19130
19140
19150
19160
19170
19180
19190
19200
19210
19220
19230
19240
19250
19260
19270
19280
19290
19300
19310
19320
19330
19340
19350
19360
19370
19380
19390
19400
19410
19420
19430
19440
19450
19460
19470
19480
19490
19500
19510
19520
19530
19540
19550
19560
19570
19580
19590
19600
19610
19620
19630
19640
19650
19660
19670
19680
19690
19700
19710
19720
19730
19740
19750
19760
19770
19780
19790
19800
19810
19820
19830
19840
19850
19860
19870
19880
19890
19900
19910
19920
19930
19940
19950
19960
19970
19980
19990
20000
20010
20020
20030
20040
20050
20060
20070
20080
20090
20100
20110
20120
20130
20140
20150
20160
20170
20180
20190
20200
20210
20220
20230
20240
20250
20260
20270
20280
20290
20300
20310
20320
20330
20340
20350
20360
20370
20380
20390
20400
20410
20420
20430
20440
20450
20460
20470
20480
20490
20500
20510
20520
20530
20540
20550
20560
20570
20580
20590
20600
20610
20620
20630
20640
20650
20660
20670
20680
20690
20700
20710
20720
20730
20740
20750
20760
20770
20780
20790
20800
20810
20820
20830
20840
20850
20860
20870
20880
20890
20900
20910
20920
20930
20940
20950
20960
20970
20980
20990
21000
21010
21020
21030
21040
21050
21060
21070
21080
21090
21100
21110
21120
21130
21140
21150
21160
21170
21180
21190
21200
21210
21220
21230
21240
21250
21260
21270
21280
21290
21300
21310
21320
21330
21340
21350
21360
21370
21380
21390
21400
21410
21420
21430
21440
21450
21460
21470
21480
21490
21500
21510
21520
21530
21540
21550
21560
21570
21580
21590
21600
21610
21620
21630
21640
21650
21660
21670
21680
21690
21700
21710
21720
21730
21740
21750
21760
21770
21780
21790
21800
21810
21820
21830
21840
21850
21860
21870
21880
21890
21900
21910
21920
21930
21940
21950
21960
21970
21980
21990
22000
22010
2202
```

231 240 245 246 250 253 260  
val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TCC TGC CAT GGA GAT CTG CTT GAA TGT CCT GAT CAC ACG GCG GAC CTT (890)

261 265 270 278 279 280 289 290  
ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys ile  
GCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TCT CAC TCC ATT (980)

291 300 310 316 320  
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
CCC GAA GTG GAA AAT GAT CAG ATG CCT GCT GAC TTG CCT TCA TTA GCT CCT GAT TTT GTT GAA AGT AAG GAT GTT TGC AAA AAC TAT GCT (1070)

321 330 340 350  
glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala  
CAG GCA AAG CAT GTC TTC TTG CCC ATG TTT TTG TAT CAA TAT GCA AGA AGG CAT CCT GAT TAC TCT GTC CTG CTG ACA CTT CCC (1160)

351 360 361 369 370 380  
lys thr tyr glu thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
AAG ACA TAT CAA ACC ACT CTA GAG AAG TGC TGT GCC GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTG TTC CAT GAA TTT AAA CCT CCT (1250)

381 390 392 400 410  
val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu phe glu asn ala leu leu val arg  
GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CAG CTT GCA GAG TAC AAA TTC CAG AAT CCG CTG TTA GTT CTT (1360)

411 420 430 437 438 440  
tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gly lys val gly ser lys cys cys lys his  
TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA ACA AAC CTA GCA AAA GTG GCC ACC AAA TGT TGT AAA CAT (1430)

441 448 450 460 461 470  
pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser  
CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CAG TTA TGT GTG TTG CAT GAG AAA ACG CCA GTA AGT (1520)

471 476 477 480 490 500  
asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
CAC ACA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG CCA CCA TCC TTT TCA GCT CTG GAA GTC CAT GAA ACA TAC GTT CCC AAA (1610)

501 510 514 520 530  
glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg gln ile lys lys gln thr ala leu val  
CAG TTT AAT GCT GAA ACA TTC ACC TTC CAT CCA GAT ATA TCC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAC AAA CAA ACT GCA CTT GTT (1700)

ter

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

Met lys trp val tlu phe ile ser leu leu phe leu phe ser  
ATG AAG TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT AGC (30)

GCYTTTCTCTCTGTCAACCCACACGCCCTTTGGCACA

-18 p r o -10

-1 -6 -1

ser ala tyr ser arg gly val phe arg arg  
TCC GGT TAT TCC ACC GGT GTG TTT CGT CGA

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

```

5      10      15      20      25      30      35
-6   p r o  -1   1
arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asn phe lys
ACG GGT GTG TTT CGT CCA CAT GCA CAC AAG AGT GAG GTT GCT CAT CGG TTY AAA GAT TTG GCA GAA GAA AAT TTC AAA (170)

21      30      40      50
ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GTA AAA TTA GTG AAT CAA GTA ACT CAA TTT GCA (260)

51      60      70      75      80
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
AAA ACA TGT GTT GCT GAT GAG TCA CCT CAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA GAC AAA TTA TGC ACA GTT GCA ACT CTT (350)

81      90      91      100      101
arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu gln his lys asp asp asn pro
CGT GAA ACC TAT CGT CAA ATG CCT CAC TGC TGT GCA AAA CAA GAA CCT GCG AGA AAT GAA TGC TTC CAA CAC AAA GAT GAC AAC CCA (440)

111      120      124      130
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try
AAC CTC CCC CGA TTG GTG ACA CCA CAG GTT GAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT (530)

141      150      160      168      169      170
glu ile ala arg arg his pro tyr phe phe ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu cys cys gln
CAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG CAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA CAA TGT TGC CAA (620)

171      177      180      190      200
ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg leu lys cys
GCT GCT CAT AAA GCT GCC TGC CTG TTG CCA AAG CTC GAT GAA CTT CCG GAT CAA GCG AAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (710)

201      210      220      230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu
GCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA GCA TGG GCA GTA GCT CGC CTG AGC CAG ACA TTT CCC AAA GCT GAG TTT GCA CAA (800)

```

231 240 245 246 250 253 260  
val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
GTT TCC AAG TTA GTG ACA CAT CTT ACC AAA GTC CAC ACG GAA TGC TCC CAT GGA CAT CTG CTT GAA TGT GCT CAT GAC ACG GCG GAC CTT (890)

261 265 270 278 279 280 289 290  
ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile  
CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GAA AAA CCT CTG TTC GAA AAA TCC CAC TGC ATT (980)

291 300 310 316 320  
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
GCC GAA GTG GAA AAT GAT CAG ATG CCT GCT GAC TTG CCT TCA TTA CCT GCT CAT TTT GTT GAA AGT AAC GAT GTT TGC AAA AAC TAT GCT (1070)

321 330 340 350  
glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala  
GAG GCA AAG CAT GTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA ACA ACG CAT CCT CAT TAC TCT GTC CTG CTG GCA CTT GCC (1160)

351 360 361 369 370 380  
lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
AAG ACA TAT GAA ACC ACT CTA GAG AAG TGC TGT GCT GCT GCA CAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)

381 390 392 400 410  
val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu qly glu tyr lys phe gln asn ala leu leu val arg  
GTG GAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CGT (1360)

411 420 430 437 438 440  
tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his  
TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA GGA AAA GTG GCG AGC AAA TGT TGT AAA CAT (1430)

441 448 450 460 461 470  
pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser  
CCT GAA CCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTC CAT CAG AAA ACG CCA GTA AGT (1520)

471 476 477 480 500  
asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
GAC ACA GTC ACC AAA TCC TGC ACA GAA TCC TTG GTG AAC AGG CCA CCA TCC TTT TCA GCT CTG CAA GTC CAT CAA ACA TAC GTT CCC AAA (1610)

501 510 514 520 530  
glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg gln ile lys lys gln thr ala leu val  
GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT CCA CAT ATA TCC ACA CTT TCT GAG AAG GAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

5  
10  
15  
20  
25  
30  
35

531 glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys oys cys lys  
CAG CTC GTG AAA CAC AAG CCC AAG GCA ACA AAG CAG CAA CTG AAA CCT GTT ATC CAT TTT GCT GCT TTT GTA GAG AAG TGC TGC AAG (1790)  
558 559 560

540  
561 ala asp asp lys glu thr oys phe ala glu gln lys lys leu val ala ala ser gln ala ala leu qly leu ter  
GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG CAG GGT AAA AAA CTT GTT CCT GCA AGT CAA CCT GCC TTA TAA CATCACATTTTAAAG (1883)  
567 570 580

580  
581 ter ter  
CATCTCAGCCTACCATGAGATTAAGAGAGAAATGAGATCAAAAGCTTATTCATCTGTTTTTTTCTGTTGCTGTAAGCCACACCCCTGCTAATAAATTTCTTTAA (2002)  
TCATTTGCCCTCTTTCTCTGCTGCTTCAATTAATAAATAATGAGAAATCTAA..... 20 .....AA (2078)

[illegible]



231 val ser lys leu val thr asp leu thr lys val his thr glu cys oys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TGC TCC CAT GCA GAT CTG CTT GAA TGT GCT GAT CAC AGG GCG CAC CTT (R90)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys qiu lys pro leu leu qiu lys ser his cys ile  
CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GAA AAA CCT CTG TTC GAA AAA TCC CAC TCC ATT (R90)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
GCC GAA GTG GAA AAT GAT CAG ATG CCT GCT GAC TTG CCT TCA TTA CCT GCT CAT TTT GTT GAA AGT AAG GAT GTT TGC AAA AAC TAT GCT (R1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala  
CAG GCA AAG GAT GTG TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA ACA AAG CAT CCT GAT TAC TCT GTC GTG CTG CAC CTT GCC (R1160)

351 lys thr tyr glu thr thr leu glu lys oys cys ala ala asp pro his qiu cys tyr ala lys val phe asp qiu phe lys pro leu  
AAG ACA TAT GAA ACC ACT CTA GAG AAG TGC TGT GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (R1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe qiu qiu leu qly glu tyr lys phe gln asn ala leu leu val arg  
GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CGT (R1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his  
TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA ACA AAC CTA GGA AAA GTG GCG AGC AAA TGT TGT AAA CAT (R1430)

441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his qiu lys thr pro val ser  
CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTG CAT GAG AAA ACC CCA GTA AGT (R1520)

471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro oys phe ser ala leu glu val asp qiu thr tyr val pro lys  
GAC AGA GTC ACC AAA TGC TGC ACA GAA TCC TTG GTG AAC AGG CCA CCA TCC TTT TCA CCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (R1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qiu ile lys lys qiu thr ala leu val  
GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT CAG AAG CAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (R1700)

5  
10  
15  
20  
25  
30  
35

531  
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys cys lys  
GAG CTC CTC AAA CAC AAG CCC AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT CAT TTC GCT TTT GTA GAG AAG TGC AAG (1790)  
558 559 560

561  
ala asp asp lys glu thr cys phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter  
GCT CAC CAT AAG CAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAG (1883)  
580

540  
CATCTCAGCCTACCATGAGATAAGACGAAAGAAATGAAGATCAAGCTTATTCATCTGTTTTCTTTTTTCGTTGGTGAAGCGCACCCCTGCTAAAAACATAAATTTCTTTAA (2002)  
567 570  
TCATTTTGGCTCTTTTCTCTGTGCTTCAATTATAAAAAATGAAAGAAATCTAA..... 20 .....AA (2078)

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13,  
10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of  
15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

0 1 2 3 4 5 6 7 8 9 1.0 1.1 1.2 1.3 1.4 15 16 17 18 19 20

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**